Argentina Real Estate Listings

Author: Dharani Doppalapudi, Saiabhinav Chekka

Introduction

This dataset provides rent and sales information from 2016 for homes in Argentina. The data branch of the Latin American real estate search engine Properati is called Properati Data. This dataset includes ad_type, id, start_date, end_date, created_on, l1, l2, property_type, operation_type. Notice start date is represented by start_date, notice termination date is represented by end_date, release date of the initial version of the notice is indicated by created_on, operation_type represents if the operation is rent or sale, property_type represents if the property is apartment, ph or house. Property located in which country is given by l1 and the province details are given by l2, id is the notice identifier, ad_types gives the information of if it’s a property or project. This is a large dataset with one million rows.

Background

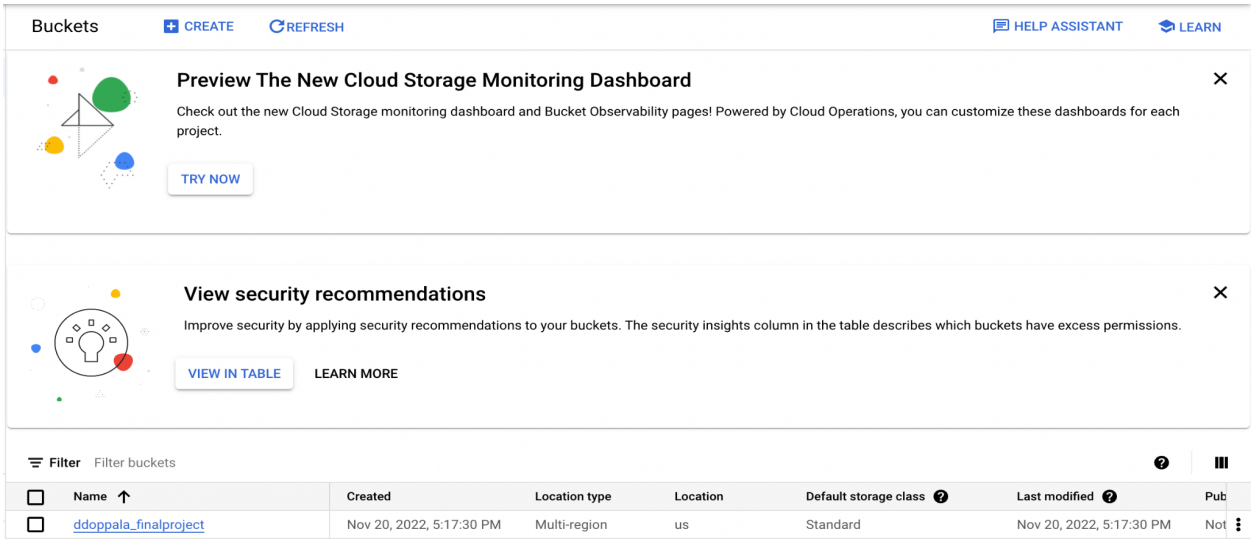
I chose Argentina Real Estate Listings for my project because it contains a large amount of data and I've always been interested in assisting Real Estates. It is important to address this because it helps the developers in detailing their properties and describing them easily and helps them sell more properties. It will help the customers in listing the available properties for sale or rent very easily and find the properties that fit their needs and budgets and the optimum times to purchase them.

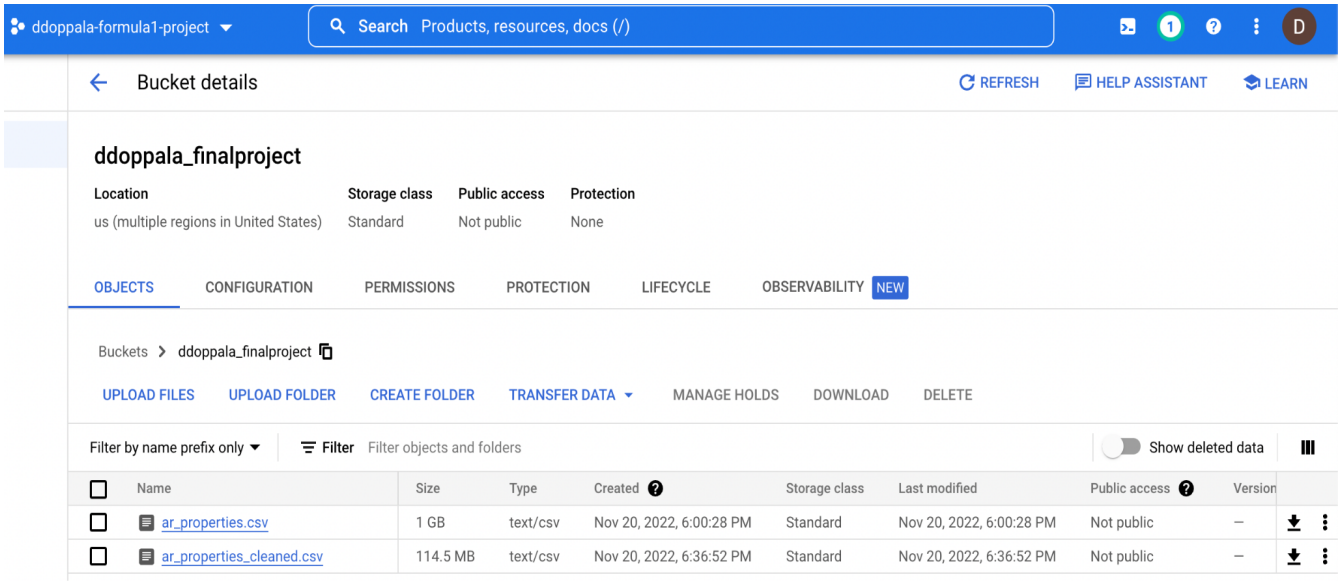
Methodology

To address this problem, I have selected the pipeline approach. I implemented a pipeline by following the steps of downloading the data, transforming the data, summarizing the data and visualizing the results and representing them in an understandable manner. Additionally, this project allowed me the chance to practice data cleaning on actual data.

I downloaded the Argentina Real Estate Listings data from the recommended Big Query datasets and after downloading, I have used Google Cloud Platform and created a bucket with the name ddoppala_finalproject in the Google Cloud Storage. Initially I uploaded the csv file of the Argentina Real Estate Listings dataset. Then I performed data cleaning and I uploaded the cleansed and transformed data csv file to my bucket with the name ar_properties_cleaned.csv.

This is my final cleaned data file where I performed a few queries to understand and represent the relationship between different fields.





Downloaded the data and explored it, and performed data cleaning by removing the missing values and transformed this cleaned data to another csv file and used this cleaned csv file for further analysis.

```
[ ] import pandas as pd
```

```
[ ] from google.colab import drive
    drive.mount('/content/drive')
```

```
Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount("/content/drive", force_remount=
```

```
[ ] dataframe = pd.read_csv("/content/drive/MyDrive/ar_properties.csv")
    print(dataframe.shape)
```

```
(1000000, 25)
```

```
[ ] dataframe.head(10)
```

	id	ad_type	start_date	end_date	created_on	lat	lon	l1	l2	l3	...
0	DyVXfkpKygVBKuUk5olH+A==	Propiedad	2020-08-22	2020-09-03	2020-08-22	-34.407468	-58.957367	Argentina	Bs.As. G.B.A. Zona Norte	Pilar	...
1	9naojilaMecJN4jlQiTkGg==	Propiedad	2020-08-22	2020-09-04	2020-08-22	-37.996039	-57.542509	Argentina	Buenos Aires Costa Atlántica	Mar del Plata	...
2	tlCPRJPjoDEUzuuCelemAQ==	Propiedad	2020-08-22	2020-08-31	2020-08-22	-31.380187	-58.009182	Argentina	Entre Ríos	Concordia	...
3	Zw3b91glQUO3HNrM5fPYlQ==	Propiedad	2020-08-22	2020-09-04	2020-08-22	-27.494106	-55.123455	Argentina	Misiones	Oberá	...
4	bsU81gm9JEgtZCbTYgvykg==	Propiedad	2020-08-22	2020-09-04	2020-08-22	-32.948856	-60.630464	Argentina	Santa Fe	Rosario	...
5	3Cips/GEPuJqktYycSALCQ==	Propiedad	2020-08-22	2020-09-04	2020-08-22	-32.884278	-60.710901	Argentina	Santa Fe	Rosario	...
6	6TO5dWwmivuJ67zkztUzCA==	Propiedad	2020-08-22	2020-09-04	2020-08-22	-34.604195	-58.559827	Argentina	Bs.As. G.B.A. Zona Oeste	Tres de Febrero	...
7	VlE71iojfIFoH4piJfFQ/g==	Propiedad	2020-08-22	2020-09-04	2020-08-22	-32.948834	-60.630169	Argentina	Santa Fe	Rosario	...

```
null_val = dataframe.isnull().sum()
print(null_val)
```

```
id                        0
ad_type                   0
start_date                0
end_date                  0
created_on                0
lat                  105767
lon                  105809
l1                        0
l2                        0
l3                    34727
l4                   693838
l5                   994470
l6                  1000000
rooms                285821
bedrooms             350067
bathrooms            234878
surface_total        522169
surface_covered      512244
price                 41757
currency              44509
price_period         570130
title                     1
description              42
property_type             0
operation_type            0
dtype: int64
```

```
[ ] cleaned_dataframe = dataframe.dropna(axis=1)
    cleaned_dataframe.shape
```

```
(1000000, 9)
```

```
cleaned_dataframe.head(10)
```

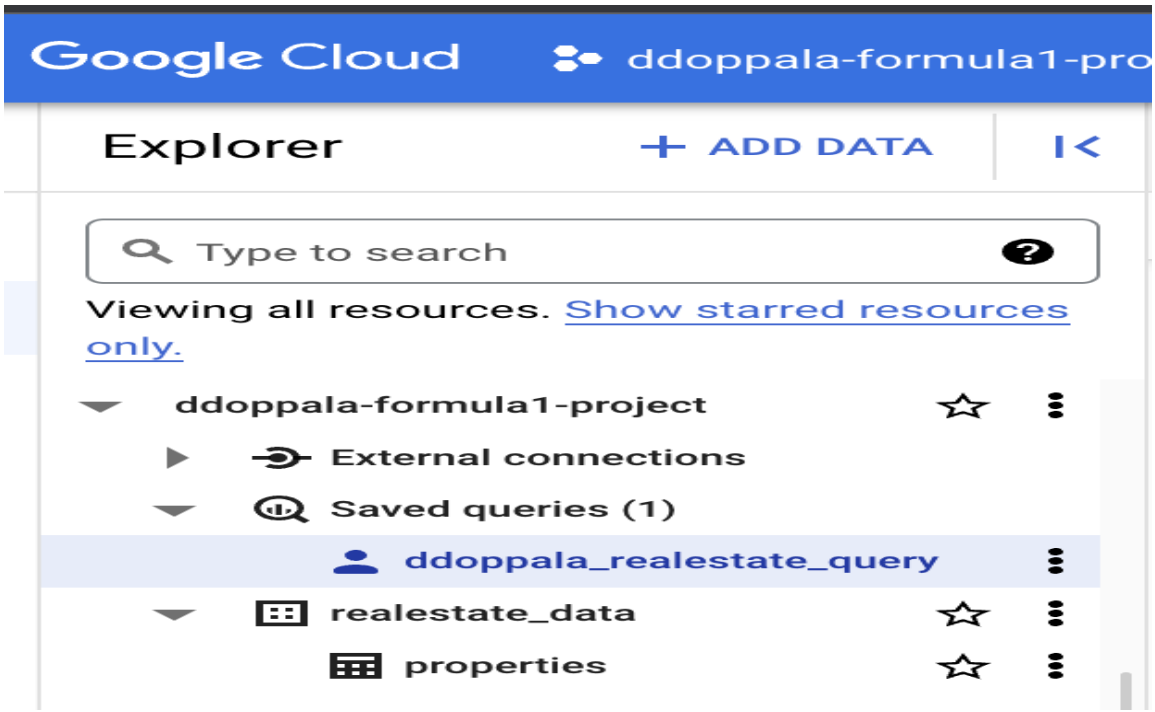
	id	ad_type	start_date	end_date	created_on	l1	l2	property_type	operation_type
0	DyVXfkpKygVBKuUk5olH+A==	Propiedad	2020-08-22	2020-09-03	2020-08-22	Argentina	Bs.As. G.B.A. Zona Norte	Lote	Venta
1	9naojilaMecJN4jlQiTkGg==	Propiedad	2020-08-22	2020-09-04	2020-08-22	Argentina	Buenos Aires Costa Atlántica	Otro	Venta
2	tlCPRJPjoDEUzuuCelemAQ==	Propiedad	2020-08-22	2020-08-31	2020-08-22	Argentina	Entre Ríos	Casa	Alquiler
3	Zw3b91glQUO3HNrM5fPYlQ==	Propiedad	2020-08-22	2020-09-04	2020-08-22	Argentina	Misiones	Lote	Venta
4	bsU81gm9JEgtZCbTYgvykg==	Propiedad	2020-08-22	2020-09-04	2020-08-22	Argentina	Santa Fe	Departamento	Venta
5	3Cips/GEPuJqktYycSALCQ==	Propiedad	2020-08-22	2020-09-04	2020-08-22	Argentina	Santa Fe	Departamento	Alquiler
6	6TO5dWwmivuJ67zkztUzCA==	Propiedad	2020-08-22	2020-09-04	2020-08-22	Argentina	Bs.As. G.B.A. Zona Oeste	Departamento	Alquiler
7	VlE71iojfIFoH4piJfFQ/g==	Propiedad	2020-08-22	2020-09-04	2020-08-22	Argentina	Santa Fe	Departamento	Venta
8	qfqOOoyVcoIjoOzxFQc+ag==	Propiedad	2020-08-22	2020-08-25	2020-08-22	Argentina	Bs.As. G.B.A. Zona Norte	Casa	Venta
9	RNDv8Sa+orCFXZElZ8r10w==	Propiedad	2020-08-22	2020-09-04	2020-08-22	Argentina	Santa Fe	Departamento	Venta

```
cleaned_dataframe.to_csv("ar_properties_cleaned.csv")
```

Big Query:

Created a project named ddoppala-formula1-project in the Big Query of the Google Cloud Platform. Within this project I have loaded the cleaned csv file from the ddoppala_finalproject bucket that I have created in Google Cloud Storage.

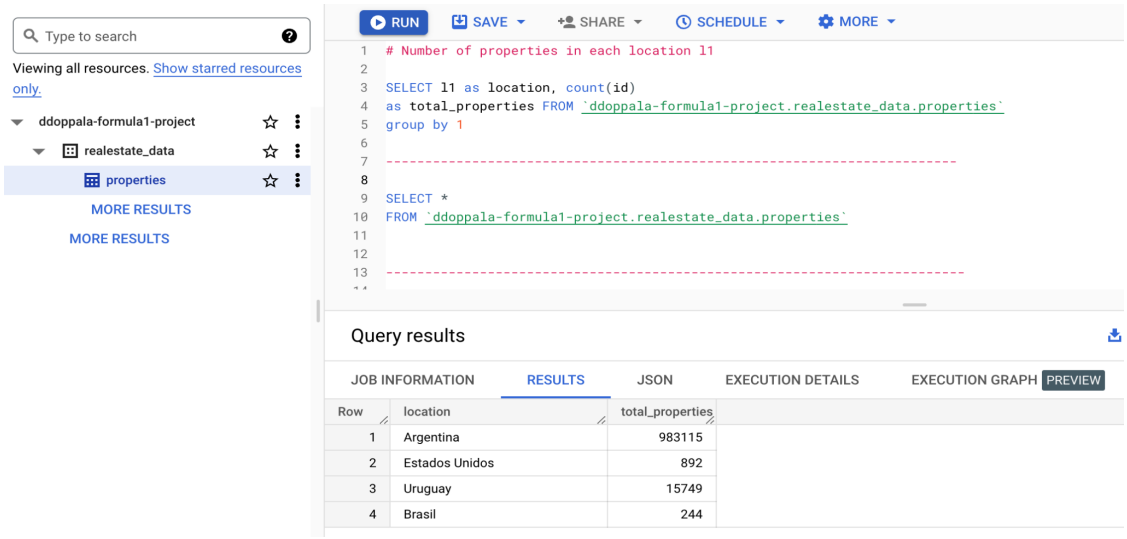
In this Big Query project, I created a realestate_data.properties table from the cleaned csv file data to perform some queries on the data. This will help in summarizing the fields of data and the relationships between them and helps in understanding the data in an efficient way. I have saved these queries in a single query file named ddoppala_realestate_query.



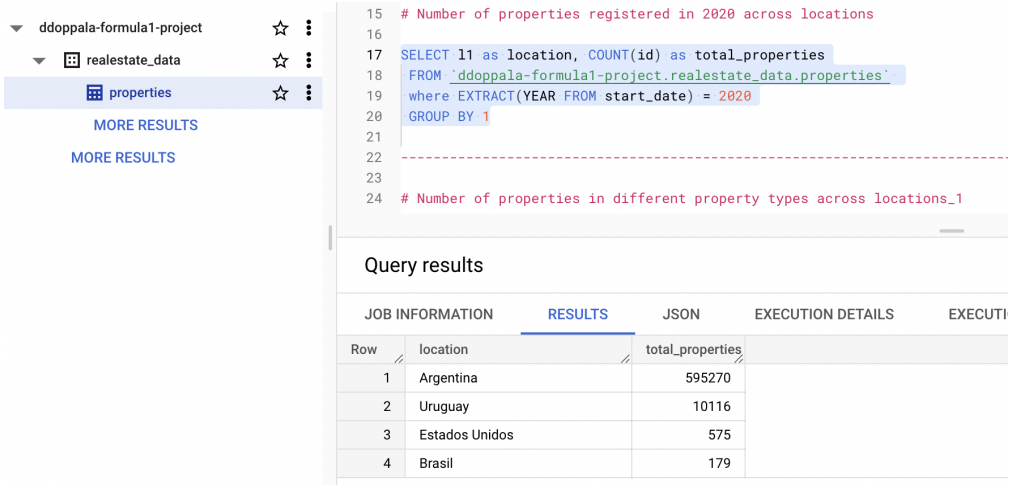
Results

One of my queries is counting the number of properties in each country. Below is the code for counting the total number of properties in each location. We can observe that there are 983115 properties in Argentina, 892 properties in the Estados Unidos, 15749 properties in Uruguay, and 244 properties in Brazil. This inquiry aids developers in knowing precisely how many of their properties are available for rent and sale in each area, as well as the count may be deducted once

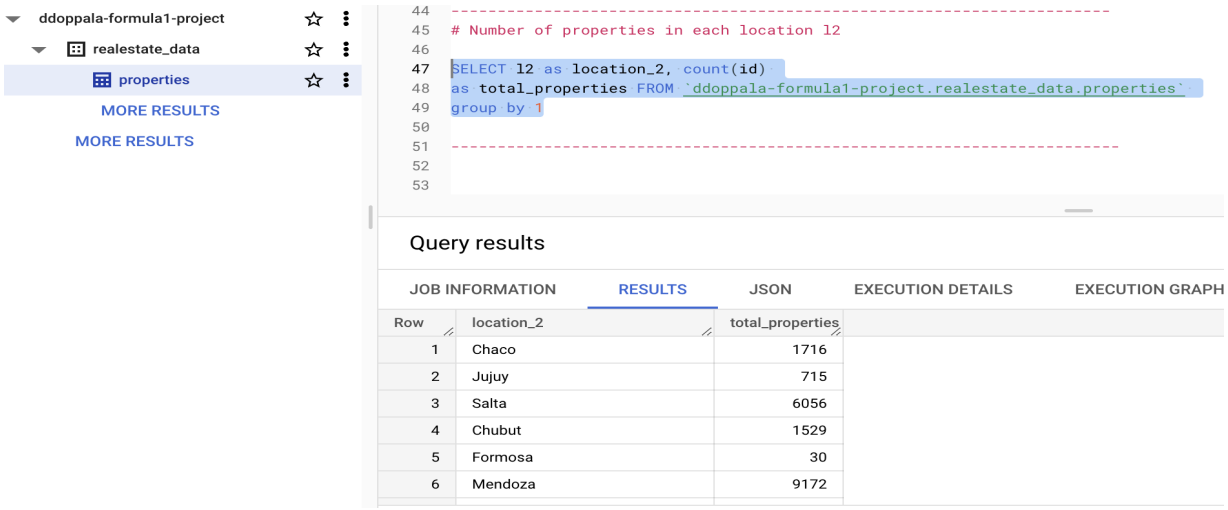
they are sold. Additionally, it is useful for clients to know which area has more properties and which has the fewest.



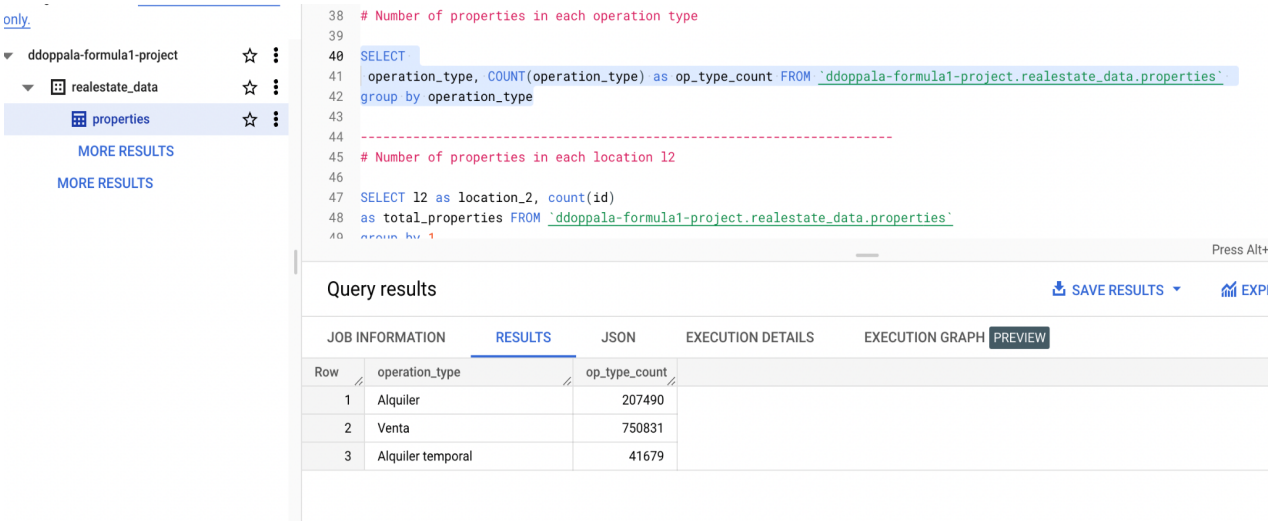
This below query calculates the number of properties registered in 2020 across locations. We can observe that there are 595270 properties in Argentina, 10116 properties in the Estados Unidos, 575 properties in Uruguay, and 179 properties in Brazil. I checked for 2020, this way we can check for any year. This helps both developers to keep track of how many properties are registered each year and helps customers to keep track of the latest registrations and the old ones and choose depending on their interests.



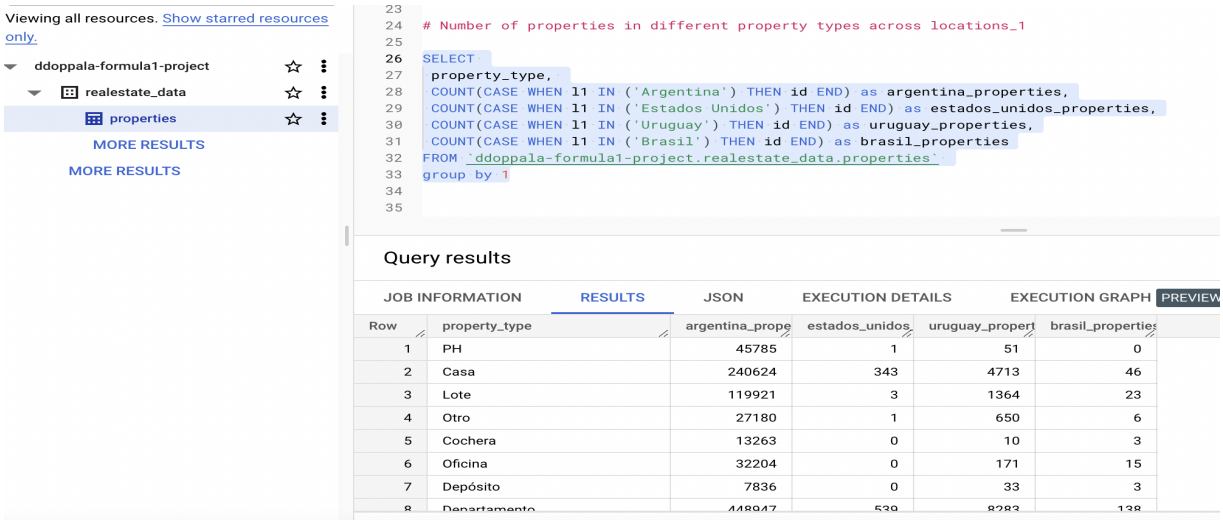
The below query calculates the number of properties in each province. As the first query this one lists the total number of properties but in each province. We can observe that there are 1716 properties in Chaco, 715 properties in Jujuy, 6056 properties in Salta, and 1529 properties in Chubut and so on. This helps the developers to keep track of the number of properties they have in each province and helps the customers to check the number of properties in the provinces they want to check and select the locations that are nearer to the places they want.



This below query counts the number of properties in each operation type. We can know the properties that are for rent or sale and other options that are available. This would be easier for customers to easily check if there are any available rent options or sale options and can check if there are other options of operations as well and if there are how many. From the below results, we could see that Alquiler operation type has 207490 properties, Venta has 750831 properties and Alquiler temporal has 41679 number of properties.

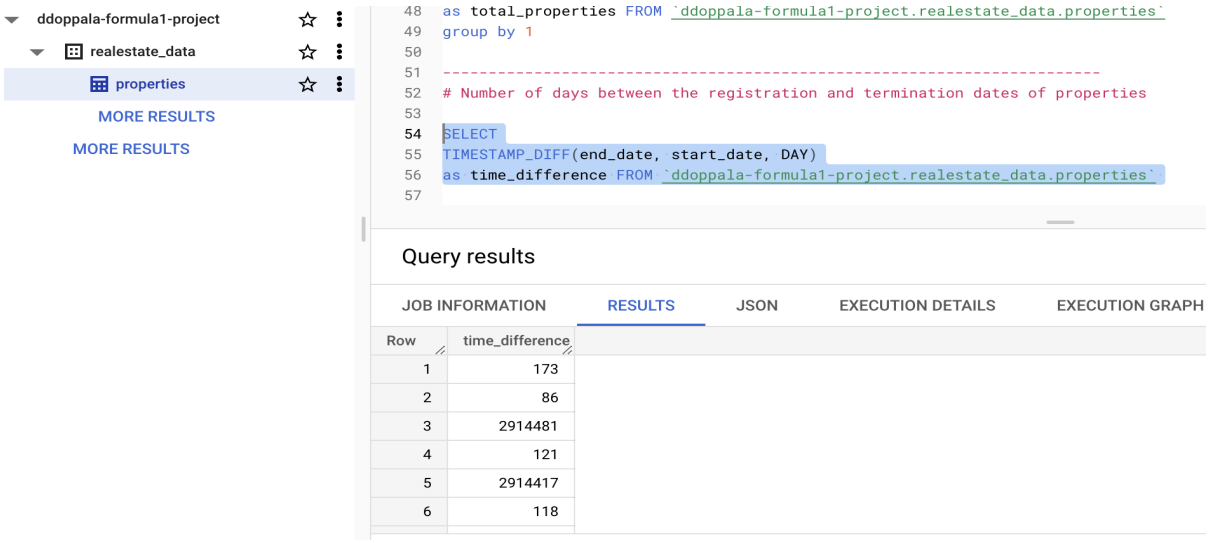


From the first query we have seen the number of properties in each country, now, in the below query we are checking the number of properties in different property types across each country. This gives us the information of the number of properties a country has and also the property type it belongs to. Gives information on how many of the properties in each country are of specific property types. In the below results, we could see that, if we select a property type PH, we could see that Argentina had 45785 properties of type PH, and Estados Unidos has 1 property of type PH and Uruguay has 51 properties of type PH and Brasil has 0 properties of type PH. In the similar way for all other property types.



Row	property_type	argentina_prope	estados_unidos	uruguay_propert	brasil_properties
1	PH	45785	1	51	0
2	Casa	240624	343	4713	46
3	Lote	119921	3	1364	23
4	Otro	27180	1	650	6
5	Cochera	13263	0	10	3
6	Oficina	32204	0	171	15
7	Depósito	7836	0	33	3
8	Departamento	448947	539	8283	138

The below query tells the information about the number of days between the day of registration and the day of termination of each property. This would help the developers to keep track of each property's duration when they are on sale. This helps the developers understand how many days each property is taking to get sold. By this we could see which properties are not getting sold for a long time and can check what steps to take to get them sold or otherwise.



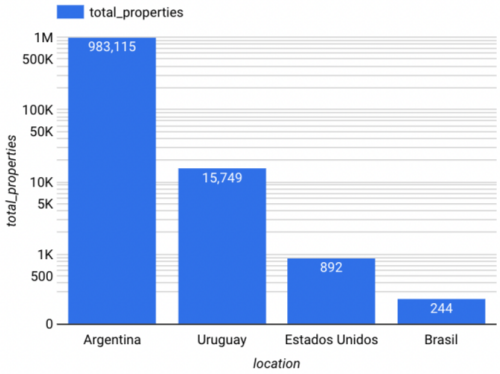
Discussion

Interpretation of the results:

This first visualization is representing the first query. It is the information about the number of properties in each country. With this, we could see that Argentina has almost all properties and just 244 in Brasil, the least number of properties. And the others are also nearer to Brasil compared to Argentina. So, we could say that the majority of the properties are in Argentina.

Number of properties in each location l1

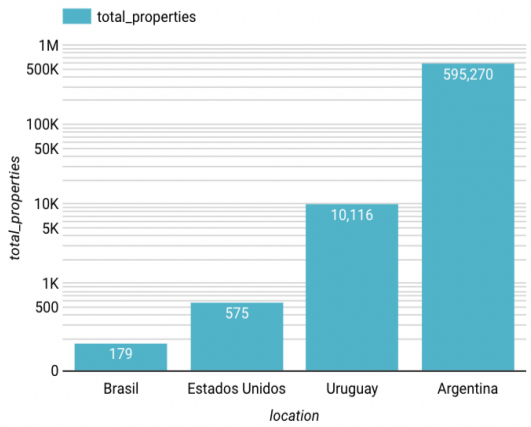
	location	total_properties
1.	Argentina	983,115
2.	Uruguay	15,749
3.	Estados Unidos	892
4.	Brasil	244



This below visualization represents the second query. It is the information on the number of properties registered in 2020 across the countries. From the below visualization we could see that the majority of the registered properties in 2020 are in Argentina of 595,270 and the very minimum of 179 are in Brasil. Uruguay has 10,116 of registered properties as of 2020 and Estados Unidos has registered 20202 properties of 575. These two countries' numbers are nearer to Brasil compared to Argentina. Argentina is dominating with the majority of the total properties registered in 2020.

Number of properties registered in 2020 across locations

	location	total_properties
1.	Argentina	595,270
2.	Uruguay	10,116
3.	Estados Unidos	575
4.	Brasil	179

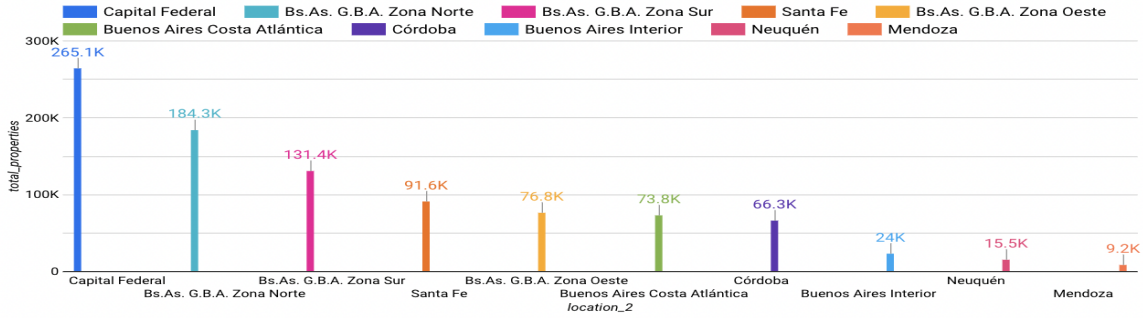


The below visualization represents the third query, calculating the number of properties in each province. From the visualization we could see that the Capital Federal has major properties of around 265.1K, and Mendoza has least of 9.2K. Bs.As.G.B.A Zone Oeste has almost a similar number of properties as Buenos Aires Costa Atlantica, with just a difference of 3K. From the graph, we could see the descending order of the province's number of properties. The major properties provinces are towards the left and minor properties provinces are towards the right.

Number of properties in each location l2

	location_2	total_properties
1.	Capital Federal	265,125
2.	Bs.As. G.B.A. Zona Norte	184,310
3.	Bs.As. G.B.A. Zona Sur	131,431
4.	Santa Fe	91,565
5.	Bs.As. G.B.A. Zona Oeste	76,786
6.	Buenos Aires Costa Atlántica	73,759
7.	Córdoba	66,274
8.	Buenos Aires Interior	23,975
9.	Neuquén	15,487
10.	Mendoza	9,172

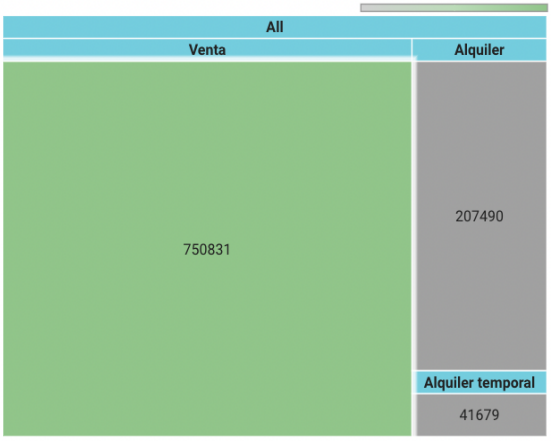
1 - 43 / 43 < >



This below visualization represents the fourth query, total number of properties in each operation type. From the below graph, we could see that there are three property types: Venta, Alquiler, Alquiler temporal. Venta has 750831 properties and Alquier has 207490 properties and Alquiler temporal has 41679 properties. From the graph, we could say that Alquier temporal has the least number of properties and Venta has the majority of the properties. Alquier has less than half the number of properties of Venta. So, Venta occupies the majority of the properties.

Number of properties in each operation type

	operation_type	op_type_count
1.	Venta	750,831
2.	Alquiler	207,490
3.	Alquiler temporal	41,679

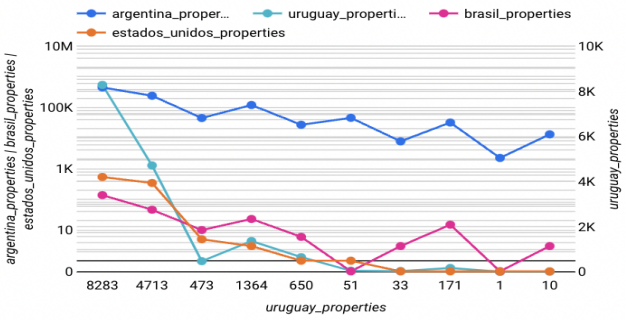


The below visualization represents the fifth query, the number of properties in different property types across locations. From the graph, we could see that the majority of properties of each property type also belong to Argentina. All other three countries' property numbers are near each other but Argentina's is far above. So, even in the different property types of each country, Argentina has a major number of properties of each property type compared to the remaining three countries. The dots in each graph, represent the property types. And for Casa de campo property type, Argentina and Uruguay have almost similar numbers of properties. Uruguay starts with major properties and the graph declines as the different property types are checked. A clear comparison of the countries and their property types is visualized below with a line graph.

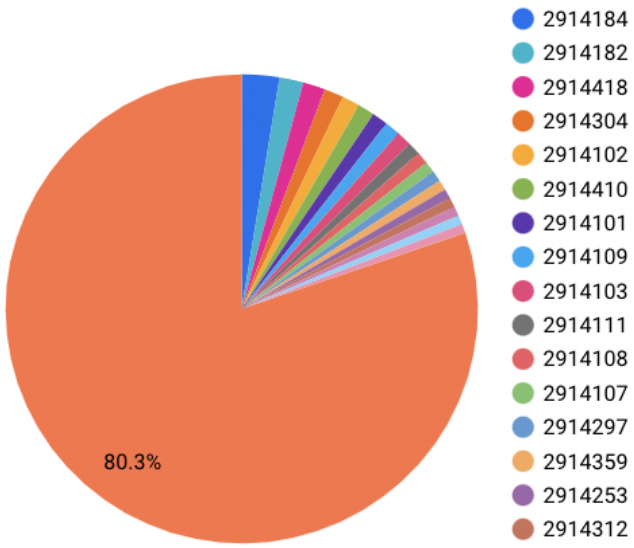
Number of properties in different property types across locations

	property_type	argentina_properties ▲
1.	Casa de campo	2,250
2.	Depósito	7,836
3.	Cochera	13,263
4.	Otro	27,180
5.	Oficina	32,204
6.	Local comercial	45,105
7.	PH	45,785
8.	Lote	119,921
9.	Casa	240,624
10	Departamento	448,947

1 - 10 / 10



This below visualization represents the last query, and tells the information about the number of days between the day of registration and the day of termination of each property. We could see the ids and percent of time taken. Most of them have taken similar time and almost 80% of the chart is occupied by id 2914304. This has taken a lot of time from the registration to the termination day. This would be considered as the oldest or the one that's been longer in sale.



Discussion of how you employed the technologies/skills from this course

I have used Google Cloud Platfrom's cloud storage to store the data and Google Cloud Platform's Big Query to access and perform query operations on data to perform data analysis and get some relations and summarizations of the data. The lectures on data pipelines, big data impact, cloud computing helped me alot to work on Cloud platform, perform queries on Big Query and in implementing pipelines. I have implemented the pipeline by downloading, data cleaning, transforming, summarizing and visualizing the data.

Any barriers or failures you have encountered:

It took me more time than I expected to perform data cleaning and transformation because the data is very large. It took me a little more time in performing some visualizations as I was trying something in Tableau and Data studio simultaneously but in the end I stuck to Data studio as I have gotten more familiar with this through our assignments.

Conclusion

It was great to apply the learning I had from the assignments and coursework in my final project. I started by cleaning and transforming the downloaded data and made a summarization on this cleaned data file. Uploaded this cleaned data file in Cloud storage and used it in my Big Query project and performed a few queries to draw some useful information from the data. Performed visualizations on these drawn results from the queries. I would like to further explore this project and draw some more insights from the dataset.

References

https://console.cloud.google.com/marketplace/product/properati/property-data-ar
https://cloud.google.com/bigquery/docs/introduction